Poisson Regression Model

Zhuoyu Jiang

12/10/2022

Question 1

In the downtown areas of very large cities, it is common for Starbucks locations to be within a block of one another. Why does Starbucks decide to put their locations so close together? One would expect that it has something to do with how busy a current location is. If an order line is long, a potential customer may not even get into line, and instead leave without making a purchase, which is lost business for the store.

Using this as motivation, a Starbucks location in downtown Lincoln, NE, was visited between 8:00a.m. and 8:30a.m. every weekday for five weeks. The number of customers waiting in line was counted at the start of each visit. The collected data are stored within the file `starbucks.csv`, where Count (number of customers) is the response variable and Day (day of week) is the explanatory variable. Using these data, complete the following.

click download *starbucks.csv*

a.

What is the population of inference? In other words, define the setting to which one would like to extend inferences based on this sample.

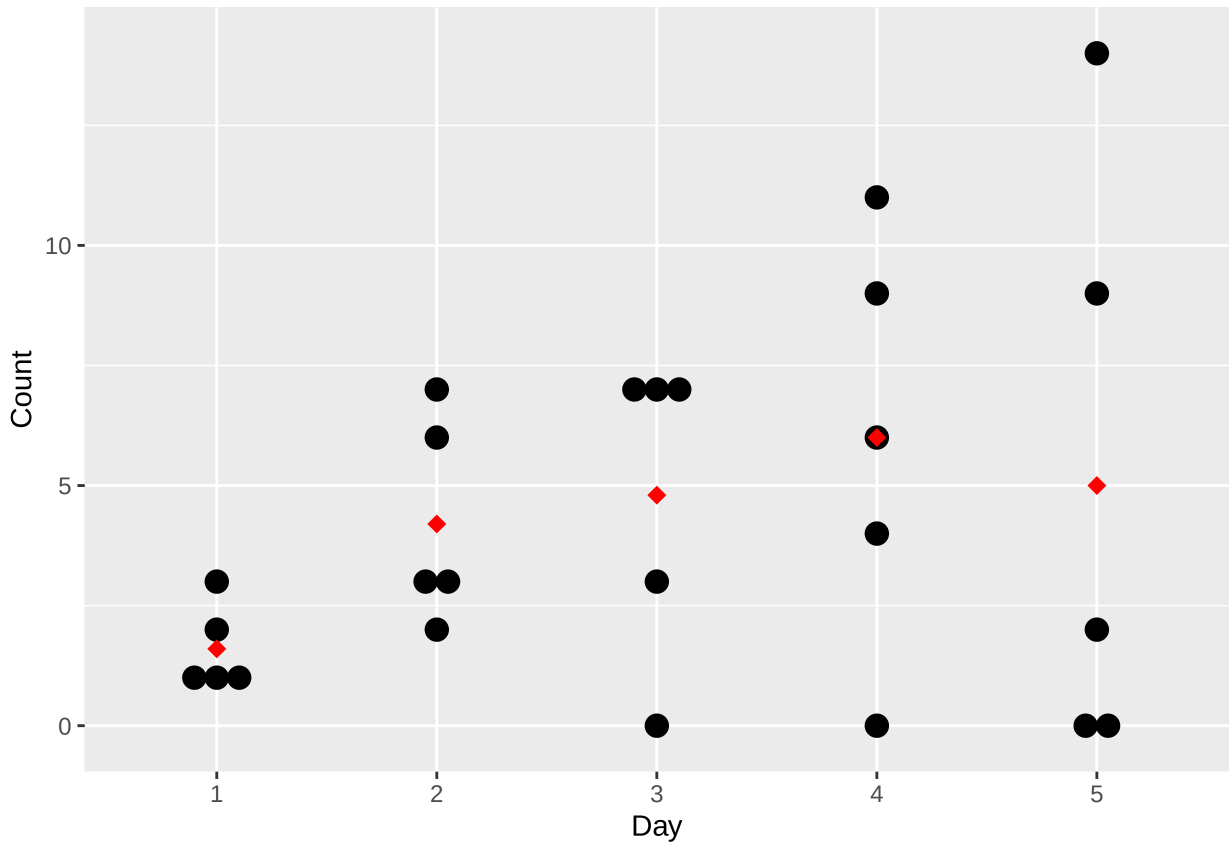
```
The population should be the number of consumers of all the starbucks stores opening during 8:00am to 8
```

b.

Construct side-by-side dot plots of the data where the y-axis gives the number of customers and the x-axis is for the day of the week. Describe what information this plot provides regarding the mean number of customers per day. In particular, does it seem plausible that the true mean count is constant across the days? (We recommend putting the factor values given within Day in their chronological order using the factor() function before completing this plot.)

```
library(readr)
starbuck <- read_csv("~/Desktop/starbucks.csv")
starbuck$Day[starbuck$Day=="Monday"]<-1
starbuck$Day[starbuck$Day=="Tuesday"]<-2
starbuck$Day[starbuck$Day=="Wednesday"]<-3
starbuck$Day[starbuck$Day=="Thursday"]<-4
starbuck$Day[starbuck$Day=="Friday"]<-5
starbuck$Day<-as.factor(starbuck$Day)
library(ggplot2)
p<-ggplot(starbuck, aes(x=Day, y=Count)) +
```

```
  geom_dotplot(binaxis='y', stackdir='center')
p + stat_summary(fun.y=mean, geom="point", shape=18,
                 size=3, color="red")
```




```
The mean number of consumer has an increasing trend on Monday to Thursday, and decreases on Friday. Fro
```

c.

Using a Poisson regression model that allows different mean counts on different days, complete the following:

i.

Estimate the model.

```
m.poisson <- glm(starbuck$Count~ starbuck$Day, family=poisson(link = "log"))
summary(m.poisson)
```

```
##
## Call:
## glm(formula = starbuck$Count ~ starbuck$Day, family = poisson(link = "log"))
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
```

```
## -3.4641  -0.8832  -0.5099   0.9392   3.2908
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)     0.4700     0.3536   1.329 0.183726
## starbuck$Day2   0.9651     0.4155   2.323 0.020188 *
## starbuck$Day3   1.0986     0.4082   2.691 0.007123 **
## starbuck$Day4   1.3218     0.3979   3.322 0.000895 ***
## starbuck$Day5   1.1394     0.4062   2.805 0.005030 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 87.432  on 24  degrees of freedom
## Residual deviance: 72.431  on 20  degrees of freedom
## AIC: 150.9
##
## Number of Fisher Scoring iterations: 5
```

The model is $logit(numall) = e^{0.47+0.9651*Tuesday+1.0986*Wednesday+1.3218*Thursday+1.1394*Friday}$

ii.

Perform a LRT to determine if there is evidence that day of the week affects the number of customers waiting in line.

H_0 : Day of week will not affect the number of customers waiting in line. H_1 : Day of week will affect the number of customers waiting in line.

```
library(car)
library(lmtest)
Anova(m.poisson)
```

```
## Analysis of Deviance Table (Type II tests)
##
## Response: starbuck$Count
##              LR Chisq Df Pr(>Chisq)
## starbuck$Day   15.002  4   0.004698 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Since the p-value is 0.004698, which is smaller than 0.05. Thus, we can reject the null hypothesis, and
```

iii.

Estimate the ratio of means comparing each pair of the days of the week and compute 95% confidence intervals for these same comparisons. Do this both with and without control for the familywise confidence level for the family of intervals. Interpret the results.

```r
library(mcprofile)
library(car)
library(multcomp)
K <- matrix ( data = c (1 , 0 , 0 , 0 , 0 ,
                        0 , 1 , 0, 0 , 0 ,
                        0 , 0 , 1, 0 , 0 ,
                        0 , 0 , 0, 1 , 0 ,
                        0 , 0 , 0, 0 , 1
                        ) , nrow = 5 , ncol = 5 ,
              byrow = TRUE )
linear.combo <- mcprofile ( object = m.poisson , CM = K )
#ci.log.mu <- confint( object = linear.combo , level = 0.95 ,adjust = "none")
#ci.log.mu
M<-contrMat(table(starbuck$Day), type = "UmbrellaWilliams")

dmcp <- mcprofile( object = m.poisson , CM = M )
summary(dmcp)
```

```
##
##    mcprofile - Multiple Testing
##
## Adjustment:   single-step
## Margin:       0
## Alternative:  two.sided
##
##              Estimate Statistic Pr(>|z|)
## C 1 == 0      0.67      0.96    0.493
## C 2 == 0      0.76      1.12    0.392
## C 3 == 0      0.72      1.06    0.428
## C 4 == 0      0.66      0.98    0.479
## C 5 == 0      0.85      1.24    0.328
## C 6 == 0      0.74      1.09    0.411
## C 7 == 0      0.66      0.97    0.484
## C 8 == 0      0.63      0.89    0.534
## C 9 == 0      0.56      0.81    0.589
## C 10 == 0     0.50      0.69    0.676
```

```r
exp(confint(dmcp)$confint)
```

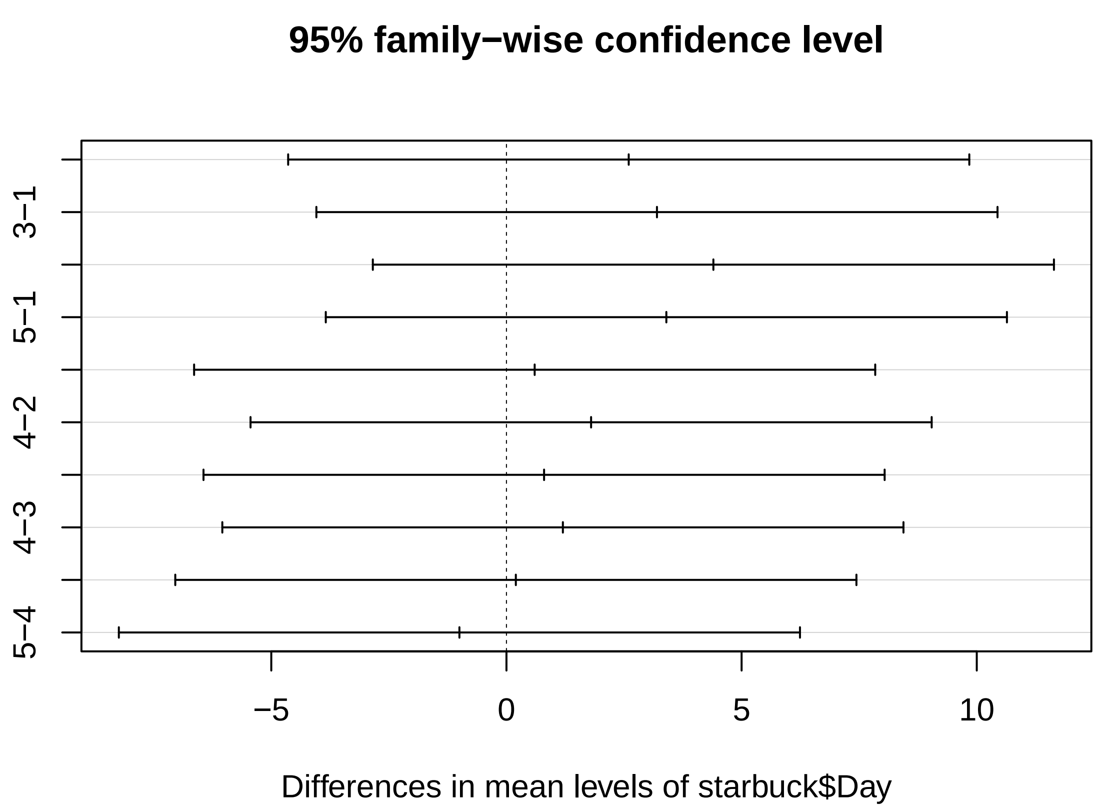
```
##        lower    upper
## 1  0.4524468 12.28689
## 2  0.5178739 13.13680
## 3  0.5011537 12.49140
## 4  0.4766467 11.78271
## 5  0.5503573 14.63604
## 6  0.5069828 12.87749
## 7  0.4720518 11.79762
## 8  0.4328737 11.81699
## 9  0.4211383 10.81799
## 10 0.3741806 10.40714
```

In the numerical output, we can find that this 95% confidence interval goes from 0.37 to 12.9 mm (lwr a

```r
# familywise-control
m.anova<-aov(starbuck$Count~ starbuck$Day)
TukeyHSD(m.anova)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = starbuck$Count ~ starbuck$Day)
##
## $`starbuck$Day`
##     diff        lwr       upr     p adj
## 2-1  2.6 -4.641299  9.841299 0.8172789
## 3-1  3.2 -4.041299 10.441299 0.6810631
## 4-1  4.4 -2.841299 11.641299 0.3911028
## 5-1  3.4 -3.841299 10.641299 0.6316343
## 3-2  0.6 -6.641299  7.841299 0.9990899
## 4-2  1.8 -5.441299  9.041299 0.9433773
## 5-2  0.8 -6.441299  8.041299 0.9971987
## 4-3  1.2 -6.041299  8.441299 0.9868380
## 5-3  0.2 -7.041299  7.441299 0.9999884
## 5-4 -1.0 -8.241299  6.241299 0.9933879
```

```r
plot(TukeyHSD(m.anova))
```



As we can see from the output above, the differences found not be statistically significant.

```
In the numerical output, we can find that this 95% family-wise confidence interval goes from -8.3 to 11
```

i.v.

Compute the estimated mean number of customers for each day of the week using the model. Compare these estimates to the observed means. Also, compute 95% confidence intervals for the mean number of customers for each day of the week.

```
day1<-exp(m.poisson$coefficients[1])
day2<-exp(m.poisson$coefficients[1]+m.poisson$coefficients[2])
day3<-exp(m.poisson$coefficients[1]+m.poisson$coefficients[3])
day4<-exp(m.poisson$coefficients[1]+m.poisson$coefficients[4])
day5<-exp(m.poisson$coefficients[1]+m.poisson$coefficients[5])

days<-c(day1,day2,day3,day4,day5)
df<-aggregate(starbuck$Count, by=list(Category=starbuck$Day), FUN=mean)

df["estimated x"]=days
df
```

```
##   Category   x estimated x
## 1        1 1.6         1.6
## 2        2 4.2         4.2
## 3        3 4.8         4.8
## 4        4 6.0         6.0
## 5        5 5.0         5.0
```

```
The estimated means and the observed means are exactly the same.
```

The 95% confidence interval is

```
confint(linear.combo)
```

```
##
##    mcprofile - Confidence Intervals
##
## level:        0.95
## adjustment:   single-step
##
##    Estimate   lower upper
## C1    0.470 -0.5012  1.20
## C2    0.965  0.0348  2.04
## C3    1.099  0.1918  2.16
## C4    1.322  0.4483  2.37
## C5    1.139  0.2392  2.20
```

d.

The hypotheses for the LRT in part (c) can be written as H_0: $\beta_2 = \beta_3 = \beta_4 = \beta_5 = 0$ VS H_A: At least one $\beta_6 = 0$. These hypotheses can be equivalently expressed as H_0: $\mu_{Monday} = \mu_{Tuesday} = \mu_{Wednesday} =$

$\mu_{Thursday} = \mu_{Friday}$ vs. H_A: At least one pair of means is unequal, where µi represents the mean number of customers in line on day i. Discuss why these two ways of writing the hypotheses are equivalent. Write out the proper forms of the Poisson regression model to support your result.

Take an example,

$$\mu_{Tuesday} = \mu_{Monday}$$

$$\frac{\mu_{Tuesday}}{\mu_{Monday}} = 1$$

$$log(\frac{\mu_{Tuesday}}{\mu_{Monday}}) = 0 = \beta_2$$

Other situations are the same.

Meanwhile,

$$\beta_2 = 0 = log(\frac{\mu_{Tuesday}}{\mu_{Monday}}) = log(\mu_{Tuesday}) - log(\mu_{Monday})$$

Thus,

$$log(\mu_{Tuesday}) = log(\mu_{Monday})$$

and

$$\mu_{Tuesday} = \mu_{Monday}$$

Question 2

click download *DeHartSimplified.csv*

Analyzing alcohol consumption as a function of number of positive and negative events for a sample of moderate-to-heavy drinkers during their first Saturday on the study.

```
DHS <- read_csv("~/Desktop/DeHartSimplified.csv")
DHS.5 <- DHS[DHS$dayweek == 5, c(1,4,7,8)]
mod.negpos <- glm( formula = numall ~ negevent*posevent , family =poisson(link = "log") , data = DHS.5 )
summary(mod.negpos)
```

```
##
## Call:
## glm(formula = numall ~ negevent * posevent, family = poisson(link = "log"),
##     data = DHS.5)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.5476  -1.3672  -0.1178   0.4927   4.9693
##
## Coefficients:
##                   Estimate Std. Error z value Pr(>|z|)
## (Intercept)         1.2214     0.1601   7.629 2.37e-14 ***
## negevent           -0.2377     0.2964  -0.802   0.4227
## posevent           -0.2300     0.1372  -1.676   0.0937 .
## negevent:posevent   0.3804     0.1863   2.042   0.0411 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
```

```
##     Null deviance: 234.89  on 88  degrees of freedom
## Residual deviance: 228.34  on 85  degrees of freedom
## AIC: 450.59
##
## Number of Fisher Scoring iterations: 5
```

The estimated model is

$$log(\hat{\mu}) = 1.2214 - 0.2377 * negavent - 0.23 * posevent + 0.3804 * negavent * posevent$$

The `negavent` and `posevent` variables are negatively correlated with the number of drinks, indicating that subjects with negative or positive events drink less than those without.

The interaction term `negavent:posevent` is positive and significant indicating that as the number of negative events improves, people will increase about 0.3804 drinks for each 1-unit change in positive events.

```
confint(mod.negpos)
```

```
##                          2.5 %     97.5 %
## (Intercept)        0.899790633 1.52770127
## negevent          -0.820033934 0.34359276
## posevent          -0.502618067 0.03549498
## negevent:posevent  0.004936037 0.73605702
```

Only the profile LR confidence interval for `negevent:posevent` from confint() is $0.0049 < \mu_3 < 0.736$, quite clearly excluding 0, which is same as the model for Saturday.

```
library(car)
Anova(mod.negpos)
```

```
## Analysis of Deviance Table (Type II tests)
##
## Response: numall
##                   LR Chisq Df Pr(>Chisq)
## negevent            2.6003  1    0.10685
## posevent            0.1061  1    0.74465
## negevent:posevent   3.9405  1    0.04714 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The p-value for β_3 is 0.04714, which is the only p-value smaller than 0.05 among three parameters and as same as the model for Saturday.

All in all, the results for the Saturday model and the results for the Friday model are similar.

Question 3

The researchers in the alcohol consumption study proposed the following hypothesis (DeHart et al., 2008, p. 529): "We hypothesized that negative interactions with romantic partners would be associated with alcohol consumption (and an increased desire to drink). We predicted that people with low trait self-esteem would drink more on days they experienced more negative relationship interactions compared with days during which they experienced fewer negative relationship interactions. The relation between drinking and negative relationship interactions should not be evident for individuals with high trait self-esteem." In DeHartSimplified.csv, trait self-esteem (a long-term view of self-worth) is measured by the variable rosn, while the measure of negative relationship interactions is nrel. Conduct an analysis to address this hypothesis, using the data for the first Saturday in the study.

H_0 : The interaction between self-esteem and relationship events has no impact on an individual's total alcoholic beverage consumption

H_1 : The interaction between self-esteem and relationship events has an impact or impacts on an individual's total alcoholic beverage consumption

```
DHS.6 <- DHS[DHS$dayweek == 6,]

#rel.freq <- table(factor(DHS.6$numall,levels=0:21))/length(DHS.6$numall)
#y <- 0:21
#prob <- round(dpois(y, mean(DHS.6$numall)), 4)
#plot(y-0.1, prob, type="h", ylab="Probability", xlab="# of Drinks",main="Observed Data vs Poisson")
model.poisson  <- glm( numall  ~  nrel * rosn ,family = poisson( link = "log" ),data= DHS.6 )
summary(model.poisson)
```

```
##
## Call:
## glm(formula = numall ~ nrel * rosn, family = poisson(link = "log"),
##     data = DHS.6)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.8324  -1.6025  -0.1471   0.5059   5.9811
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  1.32343    0.46367   2.854  0.00431 **
## nrel         1.07253    0.45716   2.346  0.01897 *
## rosn         0.01642    0.13403   0.123  0.90248
## nrel:rosn   -0.28731    0.13036  -2.204  0.02752 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 250.34  on 88  degrees of freedom
## Residual deviance: 244.30  on 85  degrees of freedom
## AIC: 507.7
##
## Number of Fisher Scoring iterations: 5
```

```
Anova(model.poisson)
```

```
## Analysis of Deviance Table (Type II tests)
##
## Response: numall
##           LR Chisq Df Pr(>Chisq)
## nrel        1.0188  1    0.31281
```

```
## rosn        0.4122  1    0.52086
## nrel:rosn   4.7191  1    0.02983 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The `nrel` variable is positively correlated and significant indicating that subjects with negative relationship events drink more than those without.

Also the interaction term `nerl:rosn` is negative and significant indicating that as self esteem improves, the drinking rate will decrease about 0.287 for each 1-unit change in self esteem. In a nutshell, since the p-value is smaller than 0.05, we can reject the null hypothesis and conclude that the interaction between self-esteem and relationship events has an impact or impacts on an individual's total alcoholic beverage consumption.

Question 4

We will use a model that regresses the number of drinks consumed (`numall`) against positive romantic-relationship events (`prel`), negative romantic-relationship events (`nrel`), age (`age`), trait (`long-term`), self-esteem (`rosn`), state (`short-term`) self-esteem (`state`), and two other variables that we will create below. We will again use Saturday data only. The negevent variable is the average of the ratings across 10 different types of "life events," one of which is romantic relationships. We want to isolate the relationship events from other events, so create a new variable, negother, as 10*`negevent-nrel`. Do the same with positive events to create the variable posother.

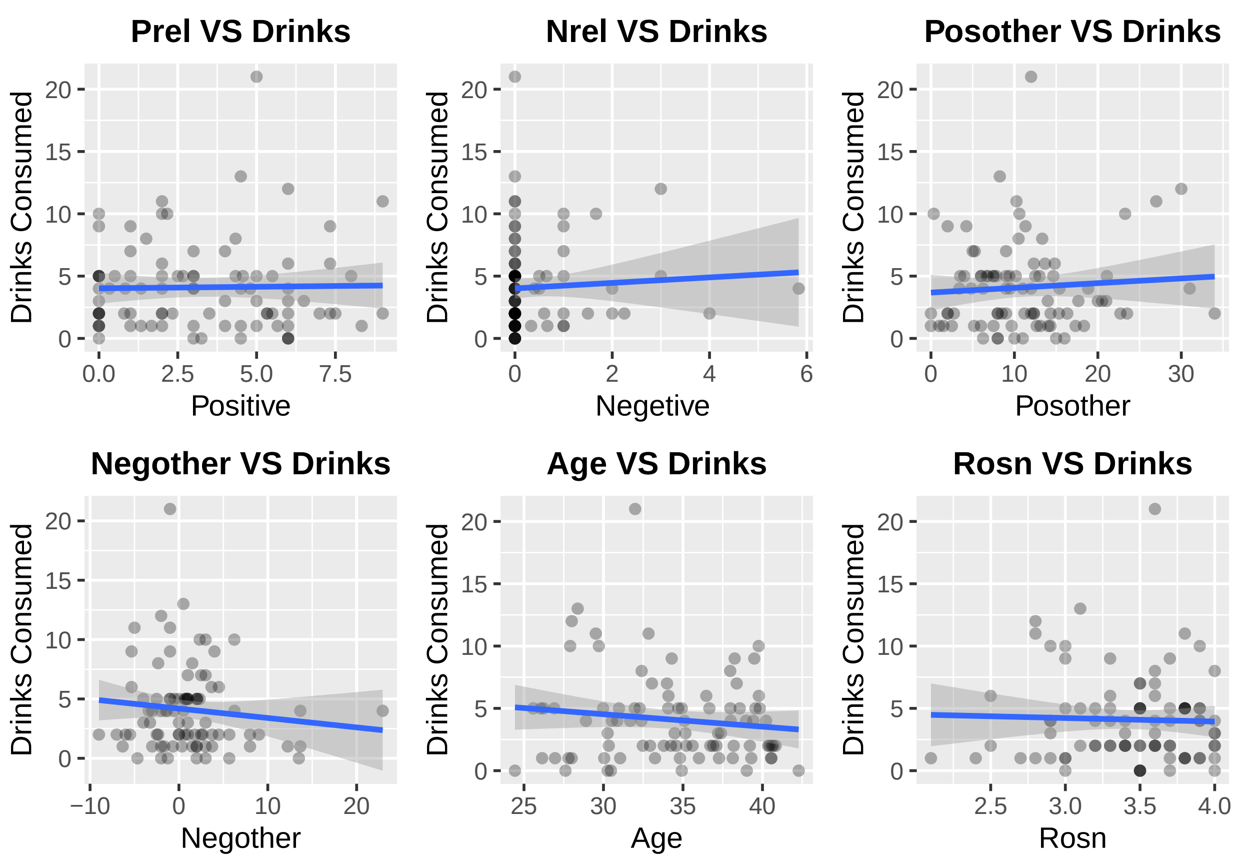
a.

Construct plots of the number of drinks consumed against the explanatory variables prel, nrel, posother, negother, age, rosn, and state. Comment on the results: which variables seem to have any relationship with the response?

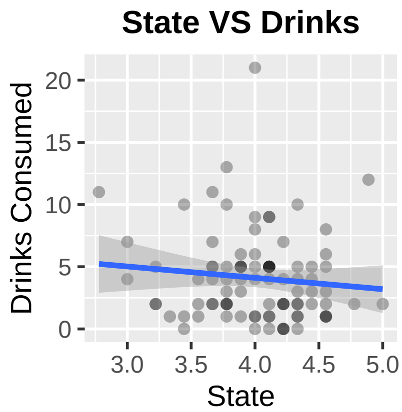
```
library(ggplot2)
library(ggpubr)
DHS.6["posother"]=10*DHS.6$posevent-DHS.6$nrel
DHS.6["negother"]=10*DHS.6$negevent-DHS.6$prel
prel_numall<-ggplot(data = DHS.6, aes(x = prel, y = numall)) +geom_point(alpha = 0.3) + geom_smooth(meth
  xlab("Positive") + ylab("Drinks Consumed") +ggtitle("Prel VS Drinks")+
  theme(plot.title=element_text(lineheight=1, face="bold", hjust = 0.5, size = 12))

neg_numall<-ggplot(data = DHS.6, aes(x = nrel, y = numall)) +
  geom_point(alpha = 0.3) + geom_smooth(method = 'lm') +
  xlab("Negetive ") + ylab("Drinks Consumed") +ggtitle("Nrel VS Drinks")+
  theme(plot.title=element_text(lineheight=1, face="bold", hjust = 0.5, size = 12))

posother_numall<-ggplot(data = DHS.6, aes(x = posother, y = numall)) +
  geom_point(alpha = 0.3) + geom_smooth(method = 'lm') +
  xlab("Posother") + ylab("Drinks Consumed") +ggtitle("Posother VS Drinks")+
  theme(plot.title=element_text(lineheight=1, face="bold", hjust = 0.5, size = 12))

negother_numall<-ggplot(data = DHS.6, aes(x = negother, y = numall)) +
  geom_point(alpha = 0.3) + geom_smooth(method = 'lm') +
  xlab("Negother") + ylab("Drinks Consumed") +ggtitle("Negother VS Drinks")+
  theme(plot.title=element_text(lineheight=1, face="bold", hjust = 0.5, size = 12))
```

```
age_numall<-ggplot(data = DHS.6, aes(x = age, y = numall)) +
  geom_point(alpha = 0.3) + geom_smooth(method = 'lm') +
  xlab("Age") + ylab("Drinks Consumed") +ggtitle("Age VS Drinks")+
  theme(plot.title=element_text(lineheight=1, face="bold", hjust = 0.5, size = 12))

rosn_numall<-ggplot(data = DHS.6, aes(x = rosn, y = numall)) +
  geom_point(alpha = 0.3) + geom_smooth(method = 'lm') +
  xlab("Rosn") + ylab("Drinks Consumed") +ggtitle("Rosn VS Drinks")+
  theme(plot.title=element_text(lineheight=1, face="bold", hjust = 0.5, size = 12))

state_numall<-ggplot(data = DHS.6, aes(x = state, y = numall)) +
  geom_point(alpha = 0.3) + geom_smooth(method = 'lm') +
  xlab("State") + ylab("Drinks Consumed") +ggtitle("State VS Drinks")+
  theme(plot.title=element_text(lineheight=1, face="bold", hjust = 0.5, size = 12))

ggarrange(prel_numall,neg_numall,posother_numall,negother_numall,age_numall,rosn_numall,state_numall, n
```

```
## $`1`
```




```
##
## $`2`
```




```
##
## attr(,"class")
## [1] "list"      "ggarrange"
```

There is not obvious correlationships between the seven explanatory factors and the response. However, the variables like `Posither`, `Negother`, `Age`, `State` may have correaltions in logistic regression model.

b.

Fit the full model with each of the variables in a linear form. Report the regression parameter estimates, standard errors, and confidence intervals. Do these estimates make sense, considering the plots from part (a)?

```
DHS.6["negother"]=10*DHS.6$negevent - DHS.6$nrel
DHS.6["posother"]=10*DHS.6$posevent -DHS.6$prel
DHS.6.2=DHS.6[,c(4,5,6,10,11,13,14,15)]
model.poisson.2 <- glm( numall  ~  . ,family = poisson( link = "log" ),data= DHS.6.2 )
summary(model.poisson.2)
```

```
##
## Call:
## glm(formula = numall ~ ., family = poisson(link = "log"), data = DHS.6.2)
##
## Deviance Residuals:
```

```
##     Min       1Q   Median       3Q      Max
## -3.1585  -1.3407  -0.4244   0.6500   5.8900
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  3.581681   0.719020   4.981 6.31e-07 ***
## nrel         0.101121   0.057422   1.761  0.07823 .
## prel         0.011789   0.021690   0.544  0.58678
## rosn         0.048692   0.130374   0.373  0.70879
## age         -0.028732   0.011811  -2.433  0.01499 *
## state       -0.344624   0.124876  -2.760  0.00578 **
## posother     0.015609   0.008967   1.741  0.08173 .
## negother    -0.053265   0.016361  -3.256  0.00113 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 250.34  on 88  degrees of freedom
## Residual deviance: 226.69  on 81  degrees of freedom
## AIC: 498.09
##
## Number of Fisher Scoring iterations: 5
```

```
confint(model.poisson.2)
```

```
##                    2.5 %       97.5 %
## (Intercept)  2.165547570  4.984851690
## nrel        -0.016058911  0.209562410
## prel        -0.030927677  0.054125591
## rosn        -0.203593157  0.307630101
## age         -0.051858694 -0.005532893
## state       -0.587319310 -0.097483731
## posother    -0.002199718  0.032978086
## negother    -0.086247618 -0.022061526
```

Explanatory factors like `nrel`, `age`, `state`, `negother` in the model shows stronger relationships with the number of drinks. The model is a Poisson regression model, which considers the special distribution of the explanatory variables, which makes sense.

c.

Conduct LRTs on the regression parameters to determine which corresponding variables make a significant contribution to the model. State the hypotheses, test statistic, p-value, and use the results to draw conclusions regarding the contributions of each variable to the model.

H_0 :All variables have no significant effect on the model. H_1 :One or more variables have a significant effect on the model.

```
Anova(model.poisson.2)
```

```
## Analysis of Deviance Table (Type II tests)
```

```
##
## Response: numall
##           LR Chisq Df Pr(>Chisq)
## nrel        2.8925  1   0.088992 .
## prel        0.2947  1   0.587249
## rosn        0.1402  1   0.708103
## age         5.8841  1   0.015278 *
## state       7.4096  1   0.006488 **
## posother    2.9612  1   0.085285 .
## negother   11.6415  1   0.000645 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The p-values of variables `age`, `state`, `negother` are 0.008809, 0.048860, 0.015278,0.006488,0.000645, which are all smaller than 0.05. Thus, we can reject the null hypothesis and conclude that variables `nrel`, `prel`, `age`, `state`, `negother` have significant effects on the model.

d.

Determine whether any variables except the two negative events variables are needed in the model. To do this, refit the model with only `nrel` and `negother`. Perform a LRT comparing the full model above with this reduced model. State the hypotheses, test statistic and df, p-value, and conclusions.

```
refit.model<-glm( numall  ~  nrel + negother ,family = poisson( link = "log" ),data= DHS.6.2 )
refit.model.2<-glm( numall  ~  prel + posother +age+rosn+state,family = poisson( link = "log" ),data= DH
refit.model.3<-glm( numall  ~  prel + age+state,family = poisson( link = "log" ),data= DHS.6.2 )
lrtest(model.poisson.2, refit.model.2)
```

```
## Likelihood ratio test
##
## Model 1: numall ~ nrel + prel + rosn + age + state + posother + negother
## Model 2: numall ~ prel + posother + age + rosn + state
##   #Df  LogLik Df  Chisq Pr(>Chisq)
## 1   8 -241.04
## 2   6 -247.28 -2 12.467   0.001963 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
lrtest(model.poisson.2, refit.model.3)
```

```
## Likelihood ratio test
##
## Model 1: numall ~ nrel + prel + rosn + age + state + posother + negother
## Model 2: numall ~ prel + age + state
##   #Df  LogLik Df  Chisq Pr(>Chisq)
## 1   8 -241.04
## 2   4 -249.28 -4 16.459   0.002461 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

As the Anova analysis in question(c), other variables like `state`, `age`, `prel` are likely to be needed in model.

H_0 : variables like `state`, `age`, `prel` are not likely to be needed in model. H_1 : variables like `state`, `age`, `prel` are likely to be needed in model.

As the LRTs results, the p-value of models with variables like `state`, `age`, `prel` are 0.001421 and 0.002461, which are both smaller than 0.05. Thus, we can reject the null hypothesis and conclude that the variables like `state`, `age`, `prel` are likely to be needed in model.